CLAIMS

We Claim:

1. A method of performing automatic speech recognition in a variable background noise environment, the method comprising the steps of:

processing a first portion of an audio signal to obtain a first characterization of the first portion of the audio signal;

comparing the first characterization to a set of reference characterizations to determine a particular reference characterization among the set of reference characterizations that most closely matches the first characterization;

updating the particular reference characterization so that the particular reference characterization more closely resembles the first characterization.

2. The method according to claim 1 further comprising the step of:

detecting an inter sentence pause; and

in response to the step of detecting, performing the step of processing the first portion of the audio signal wherein the first portion of the audio signal is included in the inter sentence pause.

3. The method according to claim 2 wherein:

the step of processing the first portion of the audio signal to obtain a first characterization includes a sub-step of:

processing the first portion of the audio signal to obtain a first set of numbers that characterize the first portion of the audio signal; and

the step of comparing the first characterization to a set of reference characterizations comprises the sub-steps of:

comparing the first set numbers to a plurality of reference sets of numbers to determining a particular set of reference numbers that most closely matches the first set of numbers.

4. The method according to claim 3 wherein the step of updating the reference characterization comprises the sub-steps of:

replacing each number in the particular set of numbers with a weighted average of the number and a corresponding number in the first set of numbers.

5. The method according to claim 4 wherein the step of comparing the first characterization to a set of reference characterizations comprises the sub-steps of:

taking a dot product between the first set of numbers and each of the plurality of reference sets of numbers.

6. The method according to claim 5 wherein:

the plurality of reference sets of numbers characterize a plurality of types of non speech audio.

7. The method according to claim 6 wherein

the plurality of reference sets of numbers are means of components of Gaussian mixtures that characterize the probability of an underlying state of a hidden markov model of the audio signal, given the first set of numbers.

8. The method according to claim 7 wherein the step of processing the first portion of the audio signal to obtain the first characterization of the first portion of the audio signal comprises the sub-steps of:

a) time domain sampling the audio signal to obtain a discretized representation of the audio signal that includes a sequence of samples;

b) time domain filtering the sequence of samples to obtain a filtered sequence of samples;

c) applying a window function to successive subsets of the filtered sequence of samples to obtain a sequence of frames of windowed filtered samples;

d) transforming each of the frames of windowed filtered samples to a frequency domain to obtain a plurality of frequency components;

e) taking a plurality of weighted sums of the plurality of frequency components to obtain a plurality of bandpass filtered outputs;

f) taking the log of the magnitude of each of the bandpass filtered outputs to obtain a plurality of log magnitude bandpass filtered outputs; and

g) transforming the plurality of log magnitude bandpass filtered outputs to a time domain to obtain at least a subset of the first set of numbers.

9. The method according to claim 8 wherein the step of processing the first portion of the audio signal to obtain the first characterization of the first portion of the audio signal further comprises the sub-steps of:

repeating sub-steps (a) through (g) for two portions of the audio signal to obtain two sets of numbers; and

taking the difference between corresponding numbers in the two sets of numbers to obtain at least a subset of the first set of numbers.

10. An automated speech recognition system comprising:

an audio signal input for inputting an audio signal that includes speech and background sounds;

a feature extractor coupled to the audio signal input for receiving the audio signal and outputting characterizations of a sequence of segments of the audio signal;

a model coupled to the feature extractor, wherein the model includes a plurality of states to which characterization of the sequence of segments are applied for evaluating *a posteriori* probabilities that one or more of the plurality of states occurred;

a search engine coupled to model for finding one or more high probability sequences of the plurality of states of the model;

a detector for detecting a specific state of the audio signal and outputting a predetermined signal when the specific state is detected; and a comparer and updater coupled to the detector for receiving the predetermined signal and in response thereto updating the model so that it more closely models one or more characterizations output by the feature extractor that correspond to the specific state.

11. The automated speech recognition system according to claim 10 wherein:

the feature extractor outputs characterizations for each of a succession of frames that include feature vectors that include cepstral coefficients;

the model comprises a hidden markov model that includes a plurality of emitting states and multi component Gaussian mixtures that give the *a posteriori* probability that a given feature vector is attributable to a given emitting state;

the detector detects an absence of speech sounds by comparing a function of one more cepstral coefficients to a threshold; and

the comparer and updater determines a mean of a multi component Gaussian mixture associated with background sounds that is closest to a feature vector that characterizes the audio signal during the absence of speech sounds, and updates the mean so that it is closer to the feature vector that characterizes the audio signal during the absence of speech sounds.

12. An automated speech recognition system comprising:

an audio input for inputting an audio signal;

an analog to digital converter coupled to the audio input for sampling the audio signal and outputting a discretized audio signal; and

a microprocessor coupled to the analog to digital converter for receiving the discretized audio signal and executing a program for performing automated speech recognition, the program comprising programming instructions for:

processing a first portion of an audio signal to obtain a first characterization of the first portion of the audio signal;

comparing the first characterization to a set of reference characterizations to determine a particular reference characterization among the set of reference characterizations that most closely matches the first characterization; and

updating the particular reference characterization so that the particular reference characterization more closely resembles the first characterization;

13. A computer readable medium storing programming instructions for performing automatic speech recognition in a variable background noise environment, including programming instructions for:

processing a first portion of an audio signal to obtain a first characterization of the first portion of the audio signal;

comparing the first characterization to a set of reference characterizations to determine a particular reference characterization among the set of reference characterizations that most closely matches the first characterization;

updating the particular reference characterization so that the particular reference characterization more closely resembles the first characterization;

processing one or more additional portions of the audio signal to obtain a one or more additional characterizations that characterize the one or more additional portions of the audio signal;

comparing the one or more additional characterizations to the set of reference characterization to find reference characterizations among the set of reference characterizations that most closely matches the one or more additional characterizations.

14. The computer readable medium according to claim 13 further comprising programming instructions for:

detecting an inter sentence pause; and

in response to the step of detecting, performing the step of processing the first portion of the audio signal wherein the first portion of the audio signal is included in the inter sentence pause.

15. The computer readable medium according to claim 14 wherein:

the programming instructions for processing the first portion of the audio signal to obtain a first characterization include programming instructions for:

processing the first portion of the audio signal to obtain a first set of numbers that characterize the first portion of the audio signal; and

the programming instructions for comparing the first characterization to a set of reference characterizations comprises the programming instructions for:

comparing the first set numbers to a plurality of reference sets of numbers to determining a particular set of reference numbers that most closely matches the first set of numbers.

16. The computer readable medium according to claim 15 wherein the programming instructions for updating the reference characterization comprise programming instructions for:

replacing each number in the particular set of numbers with a weighted average of the number and a corresponding number in the first set of numbers.

17. The computer readable medium according to claim 16 wherein the programming instructions for comparing the first characterization to a set of reference characterizations comprise programming instructions for:

taking a dot product between the first set of numbers and each of the plurality of reference sets of numbers.

18. The computer readable medium according to claim 17 wherein:

the plurality of reference sets of numbers characterize a plurality of types of non voiced audio.

19. The computer readable medium according to claim 18 wherein:

the plurality of reference sets of numbers are means of components of Gaussian mixtures that characterize the probability of an underlying state of a hidden markov model of the audio signal, given the first set of numbers.

20. The computer readable medium according to claim 19 wherein the programming instructions for processing the first portion of the audio signal to obtain the first characterization of the first portion of the audio signal comprise the programming instructions for:

a) time domain sampling the audio signal to obtain a discretized representation of the audio signal that includes a sequence of samples;
b) time domain filtering the sequence of samples to obtain a filtered sequence of samples;
c) applying a window function to successive subsets of the filtered sequence of samples to obtain a sequence of frames of windowed filtered samples;
d) transforming each of the frames of windowed filtered samples to a frequency domain to obtain a plurality of frequency components;
e) taking a plurality of weighted sums of the plurality of frequency components to obtain a plurality of bandpass filtered outputs;

f) taking the log of the magnitude of each of the bandpass filtered outputs to obtain a plurality of log magnitude bandpass filtered outputs; and

g) transforming the plurality of log magnitude bandpass filtered outputs to a time domain to obtain at least a subset of the first set of numbers.

21. The computer readable medium according to claim 20 wherein the programming instructions for processing the first portion of the audio signal to obtain the first characterization of the first portion of the audio signal further comprises programming instructions for:

applying programming instructions (a) through (g) to two portions of the audio signal to obtain two sets of numbers; and

taking the difference between corresponding numbers in the two sets of numbers to obtain at least a subset of the first set of numbers.